Information Architecture with XML

A management strategy

Peter Brown
European Parliament

JOHN WILEY & SONS, LTD

Information Architecture with XML

Information Architecture with XML

A management strategy

Peter Brown
European Parliament

JOHN WILEY & SONS, LTD

Telephone (+44) 1243 779777

E-mail (for orders and customer service enquiries): cs-books@wiley.co.uk

Visit our Home Page on www.wileyeurope.com or www.wiley.com

Neither the authors nor John Wiley & Sons, Ltd accept any responsibility or liability for loss or damage occasioned to any person or property through using the material, instructions, methods or ideas contained herein, or acting or refraining from acting as a result of such use. The authors and publisher expressly disclaim all implied warranties, including merchantability or fitness for any particular purpose. There will be no duty on the authors or publisher to correct any errors or defects in the software.

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Ltd is aware of a claim, the product names appear in capital or all capital letters. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Library of Congress Cataloging-in-Publication Data (to follow)

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library
ISBN 0 471 48679 5

Typeset in Helvetica Neue and Times Roman by WordMongers Ltd, Treen, Cornwall TR19 6LG, England

Printed and bound in Great Britain by Biddles Ltd., Guildford and Kings Lynn

This book is printed on acid-free paper responsibly manufactured from sustainable forestry,
in which at least two trees are planted for each one used for paper production.

Contents

Preface

We are not responsible for the brain we are given…
But we are responsible for what we do with it.

Anon

This book started life at sea level, on the beach of a small Greek island, was written on leave, at home or away, and finished in the Swiss Alps. It was an uphill struggle in many ways unrelated to the geography, but I like to think that I've kept my feet on the ground throughout the journey.

I have been lucky in that my professional life to date has been immensely varied. Although considered a disadvantage in some disciplines, being a 'Jack of all Trades' has served me well. It has given me a wide range of skills – in politics, communications, international relations, publishing, administration, linguistics, training and of course, information technologies – and enabled me to keep an eye always on the wider canvas, beyond the speciality of the day. For once, all these valuable experiences have come together and been useful together on the same project.

Why this book?

I was inspired to write for three reasons.

Firstly, I have always maintained that XML is not a 'toy for the boys'. It is a set of strategically vital data standards, and *not* a programming language. It will allow an enterprise to manage, and extract maximum advantage from, one of its most valuable assets: information. XML's importance cannot be underestimated, and its introduction and use must be carefully and firmly managed, and not remain the exclusive preserve of the 'back room boys'.

Many enterprises are rightly worried about the risks of becoming locked into proprietary systems and of being at the mercy of a single vendor. Badly managed, XML can be worse: locking an enterprise into a home-grown proprietary standard that only a handful of programmers can fathom.

This is meant with no disrespect to programmers or analysts who want to dive in and get things done, many of whom I have admired in my years of work. It is merely my conviction that XML should only be allowed into their hands once the senior management of an enterprise has understood its implications, power and scope in all areas of their business and administration. The target audience for this book is therefore mainly management.

The book is not intended to keep the reader up to date with the latest XML specification to be wheeled out – an issue I return to in my conclusions – but rather is the fruit of a little distance and reflection. Some authorities have already suggested that as much as two thirds of the $ millions invested each year in XML modelling activities is wasted. The problem is not the technology: it is the lack of good management.

Secondly, most XML books focus on data – in the sense of structured data destined to be managed within a database. Managing text, however, has been notoriously elusive, seen by human and machine alike as little more that a sequence of characters with a beginning, an end and a lot of unpredictable characters in between. Text can be generated from standard form letters or reports, where the structure is well understood and defined, as would be used – for example – in a mail merge program. Many XML books that take this approach make major assumptions about data organization that is simply not the reality for many ‘text-centred’ organizations. My experience has been that text is far, far, more complex than this, and can rarely be understood using this simple ‘database plus form letter’ approach. For this reason, the focus of this book is strongly text and document-centred.

Thirdly, with a focus on data processing, many XML books jump immediately into databases and programming modes. The assumption seems to be that information is already well structured in the first place and that the principal function of XML is to write programs. If it comes to merely managing well-structured data, a high-quality database is probably satisfactory as it is without converting to XML. But XML is much more than this, and many potentially valuable XML-centred projects have never left the starting block due to the perception that XML is more of a hindrance than a benefit.

This book attempts to draw the mildly-curious manager – understanding that IT strategy is important but hesitant to take the plunge – into the world of the XML standards, but using the familiar environment of their day-to-day work rather than the techno-speak of IT specialists.

The information management work of the European Union (EU) institutions and some national administrations with whom I have worked provided me with some ideas for the strategy outlined in the book. Unlike many 'data rich' environments, these administrations are 'text rich': preparing texts, notably drafting legislation, *is* their business.

Although this case might be considered as unusual, it is actually very useful for a book such as this. It has offered me important insights for large scale, complex organizations. The EU's considerable language service infrastructure is unique in the world, currently inter-working among eleven languages and moving to twenty-one by 2004. The complexity of this multilingual environment is an ideal testing ground for the advantages that XML offers. The few coding examples I give, however, are fictitious and entirely my responsibility.

Why XML?

XML is not a *technology*, but rather a *standard* that serves as a powerful medium for describing, communicating and implementing a true information management strategy. It is a computing standard that does not – or should not – belong to the IT specialists, although their support is vital, as we will see. Its basic concepts – that allow you clearly to define and express business entities and terminology – are close to management concerns. Its expressiveness is of immense value to the growing field of information architecture.

As a potentially powerful and expressive management medium, therefore, XML, and the family of standards around it, is too important a business asset to be treated as merely a technical issue.

The expressiveness of XML-based systems will come from their ability to associate real meaning to simple text and thus facilitate knowledge management. If we accept that knowledge is an organized progression and aggregation of data and information, we must come to terms with what this implies, particularly as more and more data and information are available only in electronic form.

Computer programming languages put *processing* – actions – at the centre of their concerns. XML, on the other hand, is and should remain centred on what can be processed: content – *objects*. Processing is only a means to an end. An enterprise's information – the content, whether text, data or information in all its guises – is often an end in itself. It is an increasingly valuable business asset, and XML can help manage it wisely and profitably. XML is *not* a programming language: it is a powerful standard that allows you to package and label your objects in such a way that they can be processed in whatever way is most to your advantage.

Why Information Architecture?

Well-designed information systems should be like well-designed buildings. Good buildings have a strong and stable infrastructure, foundations and supporting walls. All aspects of plumbing, wiring, choice of materials, dimensions and proportions, safety and ergonomics are the fruit of centuries of architectural standards and experience. A well-planned city will ensure that all buildings exist together harmoniously and that vital networks and utilities are adequate.

An information system that is built without foundations, without attention to the choice of materials or without standards may look good on its own, but is going to be difficult to integrate in the IT 'urban landscape'.

A preoccupation with architecture and standards will ensure that whatever you choose or need to build will stand the test of time and be fully inter-operable.

Why a management strategy?

Senior managers beyond the confines of IT departments and business units are unlikely to express much, if any, interest in the complexities of computer technology. Nor should they, and with reason: it's not their job, even if they should be concerned about resources allocated to IT and how those resources are deployed. Interest in XML, on the other hand, should become as familiar as human resources management or cost-benefit analysis in the tool-kit of any manager who wants to ensure that IT use properly reflects the needs of their enterprise.

This is because XML is different. It offers a *lingua franca* not only between analysts and developers in the IT world, but also planners and decision-makers from the management world.

Deployment of XML, and the increasingly confusing series of standards that go with it, forces an enterprise to reflect upon how it manages one of its most valuable of assets: information. At the same time, deployment is not solely a matter for the 'IT people'. Indeed, the biggest danger to long-term stability in the deployment of XML in an enterprise is that it is introduced 'by stealth', and purely in the field of application development and programming. What is needed – as I will argue throughout the book –

is for its use to be studied, understood and implemented further ‘upstream’, at the heart of business management.

XML’s introduction should not be technology-driven, or discussion of its use limited to the IT community. My approach argues that all users can be drawn in and offer valuable contributions to XML’s intelligent deployment, without getting bogged down in the technical jargon.

Who should read this book

- *Project sponsors*. If you are relying on subcontracting development work, to a specialist team or beyond the enterprise, it is important to know which issues you must keep in hand, which you can safely delegate and how, and which upstream issues need addressing, before even launching into any project.
- *Middle-managers and business analysts*. If you are to avoid the disastrous problems of proprietary formats, lack of inter-operability and wasted resources that have plagued too many IT developments in the past, it is not sufficient to believe that ‘doing it in XML’ is necessarily going to make the situation any better. It is vital to see the ‘big picture’ of what the XML family of standards can offer, and ensure cross-service cooperation, strategic planning and well-argued business cases. Initial investment in developing XML-centred systems might seem alarmingly high for no obvious initial return. A clear understanding therefore of its immense power will help in value analysis, showing favorable cost/benefit ratios and short returns on investment.
- *IT Managers*. Developers adore being let loose on a new programming language and IT environment. XML is different. If not properly thought through, you are likely to be left explaining why all your different XML projects don’t and possibly can’t work together, despite all the hype. You need to understand when and how to bring senior management on board and force them to address issues regarding XML that are not technology problems, but are firmly in their realm and require their decisions.
- *Senior management*. You are likely to be excited by the idea of reduced IT costs, improved inter-operability, faster returns on investment (RoI) and application development cycles – but can the claims made for XML all be true? What are the right questions to ask, and of whom?

The structure of the book

The book presents a hands-on approach for management, an approach essential to XML's successful deployment. It introduces the key concepts of XML in terms that managers can appreciate and subsequently deploy. It proposes an approach that enables managers to outline their desires and strategy in terms that IT specialists can appreciate, while keeping a firm management hand on the tiller.

After the introductory chapter, Chapters 2 to 4 detail the roles played by information management and XML and the need for a management-driven strategy.

Chapters 5 to 9 cover the planning, development and management of an XML-centred information architecture.

Chapters 9 to 13 then look at specific areas of XML deployment.

Food for thought

Something to chew on

Throughout the book, readers are invited to chew over a number of important and interesting issues.

This 'food for thought' is precisely the sort of insight that ought to get your management juices flowing. Savor it, share it and take your time to digest it.

It is *not* the intention of this book to delve into the technical intricacies of the XML and related standards. What little code there is, is provided to give some insight into the power of XML, and as such is not intended for use 'as is'. An overview of the key XML standards of interest and concern to management is however included in the reference section.

Food, glorious food

Analogies and metaphors can be very powerful, particularly when explaining unfamiliar ideas. An ever-curious cook myself, I use analogies to food in particular, and have found a resonance for this metaphor with managers: the separation of pre-packaged information 'dishes' (documents) into their constituent ingredients is central to managing the way in which information is created, identified, processed and ultimately consumed.

When recounting a recipe, it is rare to focus on the utensils: certainly the right tools are important to get the job done. But it is the content – the ingredients – and the recipe – the processes – that are ultimately the most important considerations. You will need the tools necessary to get the job done, no more. Some will improve efficiency – occasionally drastically – but they are not going to achieve anything unless you put them to work. That is why there is little attention given in the book to tools: I have tried to give some indications of the criteria you might apply when selecting your utensils, but your choice really depends on what meals you intend to prepare.

With these thoughts in mind, it remains for me only, therefore, to wish you…

Bon appetit!

Peter Brown

Acknowledgments

I have been helped along a difficult journey by a number of people who have contributed, often unwittingly, to the book's final publication: Michael Fitzgerald and Karen Mosman, for starting me on the road and to many others for keeping me going: Christiane Bracher, Bryan and Carole McHugh, Ralph McIntyre, Nick Marsh, Emmanuel, Alexandre and Andreas Pappamikail, Jack Fairey, Dick Stevens, Selina Barnett, Mark McLeod, Andrew Carter, Maddy Scott, Anne Wrightson, Bernard Vatant, Caroline Philips, Clive Mills, Sandra Kyriacou, Dianne Dodsworth, Suzanne and Keith Sivyer, Paul Mills, Claire Bicknell, Yuri Layhe, Helen Sims, Judi Bond, John Innes, Michelle Crozier, Stacey Robbins, Steven Stingemore, Anita Best, Leonie Riches, Anne Corkhill, Melita Jones, Sigrun Hoppmann, Colin Belchamber, Anabela and Pedro Frade, Gordon Jenner, Mark Buss, Simon Binns, Zoki Tsoi, Gary Dunleavy, Craig Walters, Anna Somerford, Rio Romanucci, Stuart Pate, Stuart Harris, John Frehley, Kate King, Rebecca Allison, Victoria Daly, Thomas Kloiber and others from the 'Eastbourne Crew'.

My sincere thanks also to Gaynor Redvers for picking up the baton at Wiley halfway through the project and soldiering on, to Steve Rickaby for timely advice and consistent support through the sometimes hair-raising last lap before publication, and to Nassos at the Sportscafé in Heraklion, Crete, without whom I might have been three chapters light.

A very big 'thank you' to you all. But especially to Christiane.

1 Introduction

Our diet will change dramatically in the future, although the essential components that we need to eat in order to stay healthy remain the same

Brian J. Ford, *The Future of Food*

A culinary tale

In the 'good old days' before food labelling, sell-by dates and competitive brand promotion, you placed yourself at the mercy of your local village store manager. After the painful wait for the previous customer to bid his farewells and finally let attention turn your way, you placed your trust in the nice old guy who knew his store and his supplies. You often ended up with more than you bargained for, with a tip thrown in on the best and freshest deals of the day, a few extra ingredients to spice up that special recipe and a summary of the latest village gossip.

The intelligence of the 'system' – the management of a wide range of foods and ingredients – was human: a customer's questions dealt with personally and a cumulative knowledge of their needs and interests allowing a truly personal service to be offered.

'Caveat emptor'

The model doesn't scale well, however. Customers today want wider choice and availability, better prices and faster service, so the supermarket revolution was born. The downside for customers was the need to 'internalize' that grocer's wisdom and assess their purchases for themselves: the shelf stackers could point you to the flours but would be hard pushed to tell which one was best for waffles. Even if they did have an opinion on the matter, they probably wouldn't have been allowed to express it, for fear of being seen to promote one brand over another.

Then is there is the question of quality and trust. In many countries, it has taken major food quality and health scares to prompt public authorities to interpose themselves between producer and consumer and insist on food labelling, quality control regulations and inspection. In parallel, the growth of the fast food outlet offered ‘no-questions-asked, no-answers-given’ solutions to the busy and/or unimaginative: fast and cheap, benefiting from economies of scale and industrial-style production, as long as you accept the pre-determined and pre-packaged realization of someone else’s flight of fancy.

‘Parse the salt’

Have you ever tried to order – let alone receive – a salt-free quarter-pounder and fries? Or asked for a reassurance that the beef is hormone or BSE free? If you are lucky, you will receive a sympathetic shrug and an explanation that “I only cash and wrap” – in other words, a simple reminder that you can take it or leave it. The consumer has no control over what is delivered and little idea of what goes into it.

If customers want a varied and balanced diet, control over the ingredients and guarantees of quality, they should not rely on fast food outlets, but rather opt for the organic store, find a reputable restaurant with a patient Maitre d’, shop around or grow their own.

Food for thought

In the good old days, committing ideas to paper was labour intensive, copying messy, impracticable or expensive, and re-using typed material without retyping unheard of. The corporate collective memory of an enterprise was a rich mix of human experience and well-filed documents.

A century ago, a secretary was a highly skilled, valued and usually male employee. The rôle of the humble and now largely defunct filing clerk was key to the corporate memory, ensuring the safe deposit and retrieval of documents.

In smaller enterprises, it was the clerk that would meet information retrieval and storage demands, often with an undocumented and highly personal system. In bigger organizations, master file classification systems were developed and documented: labelling became more sophisticated, cross-referencing introduced between files and individual documents, check-in/check-out systems and routing slips introduced. It was still often the filing clerk that would add those personal touches to the corporate system to keep it in tune with personal quirks of managers. It was still *people* that moved the documents and files around. Humans commented on their contents, their whereabouts, *who* had to do *what* to *which* text. The ‘intelligence’ of the system – like the grocery store – was human.

The increased use of electronic means of information storage – whether ‘traditional’ documents, e-mail, accounts or statistics – has lead to an immeasurably larger pool from which ever-greater information-hungry customers want to draw. But electronic files don't carry the same baggage, and it is still humans that are called upon to interpret and comment…

Except that the filing clerks, the well-documented and structured file classification systems aren't there any more – hard-pressed secretaries and administrators rely on their wits and highly personalized systems to provide the intelligence sought.

Except that the secretary isn't there on the evening the manager needs to piece together information from several documents, databases and other sources, and the report is due yesterday, as usual.

What, why and 'why me?'

This book is less about the '*how?*' of XML but first and foremost the '*why?*'

We will also see what XML can be used to tackle and who should be involved and responsible.

It is not therefore primarily a book for the programmer or application developer, who wants to explore the intricacies that the XML family of standards has to offer. There are already plenty of good books on the market that do this. This is rather a book for managers, whether they be responsible for managing resources, information and data content, editorial and design policy, business rules and processes or IT infrastructure. There are two important messages for them, which can sometimes seem contradictory:

- 'Keep the toys from the boys' – ensure that development and implementation of XML-based systems are kept firmly in the hands of management and not left to the back-room boys without any framework or guidance.
- 'Resistance is futile' – as several business analysts have pointed out, doing nothing about XML is not an option: the potential benefits of XML in so many areas of an enterprise's activities mean that it is going to pop up somewhere sooner or later. As such, it is better to be prepared and have a strategy to manage a coherent implementation before things get out of hand.

It is however valuable for a programmer or developer who wants a better grasp of the wider or senior management perspective of information management, who wants to see the entire wood before examining the trees. This is sometimes difficult when you are knee-deep in parsed external entities and library callback functions but – as we will see throughout – is absolutely essential for a robust and well-designed implementation strategy.

To start with, we examine three main questions:

- *What* problems can XML address?
- *Why* is XML important?
- *Who* should be involved and responsible for XML?

Before looking at XML, however, we should start by examining and understanding some the principal concerns under the general heading of 'information management'.

Information management

Few of us are today not involved in information management to one degree or another. Whether our main concern is managing personal accounts and an address book or a multi-Terabyte data warehouse, we are all confronted by four major challenges in our increasingly 'digital everything' world:

- Information overload
- 'Digital rot'
- Content and transaction management
- Multiplicity of formats and media for the same content

Information overload

In contrast to the early days of computing, until the last decade, the cost of long-term mass storage of digital information has fallen dramatically.

In the years when digital storage space was at a premium, and programmers ten-to-a-penny, major efforts were made to optimise code and compress content. Organizations were careful about what data and text they would save and archive, as these represented considerable overhead in the IT budget.

The emphasis was on optimising code and compressing content. Such economies were considered acceptable: why, for example, express a year as four digits, when two will surely do?

That particular problem will not arise again, for either 90 years if you didn't do anything about the Millennium Bug, or another 7000 if you did but considered four digits enough. Unfortunately, there are plenty of equally costly 'semantic short-cuts' around. A high and often costly premium is still placed on brevity.

In contrast, we are now entering the age of digital everything, rendered feasible by a combination of:

- Massive increases in storage and processing capacity – a full length feature film will fit on a DVD disk, together with dubbed soundtracks and subtitles in a handful of languages. Similarly, an average home PC will now ship with a processor a thousand times more powerful than a decade ago.
- Plummeting costs for mass storage.

As well as generating and storing more information, we are moving more of it about, copying more of it to more people and as a result storing increasing numbers of copies of the same content, while retaining more drafts and versions of incomplete texts.

There has been a vast increase in the total volume of information produced and stored digitally in the world, and yet we are not – according to the report – consuming any more than years earlier.

Food for thought

Bulimia?

A total data consumption in the USA of 3,344,783 Megabytes in 1999 alone sounds impressive.

According to a study in 2000 by the School for Information Management and Systems (SIMS) of the University of California at Berkeley, however, it is not significantly more than in 1991.

Is it possible that we actually have a relatively fixed capacity to digest information, and that we have to start dieting?

If the growth in volume of source data continues, therefore, we are faced with an equally growing need to filter and select. To do this, we therefore need to be able to identify our information more easily before selecting what we want to consume.

'Digital rot'

In the days when enterprises still employed filing clerks and corporate filing systems, everyone knew that someone was looking after the files. The ubiquity of the desktop computer file system has brought that era to an abrupt and often messy end: each user has a very personal understanding of what 'correctly filed' means, and this has heralded a breakdown of corporate-wide classification and filing systems. Complacency is compounded by the belief that one can 'always do a search' to find an elusive file, or that your software will somehow manage the problem for you.

But ‘digital rot’ has taken hold. A number of studies have concluded that, far from guaranteeing the longevity of knowledge, many digital collections actually undermine it because of three dangerous assumptions.

Firstly, there is the assumption that everything is kept, that disks don’t crash, get wiped by users or reformatted by system administrators wanting to free space on a network. The mere existence of the electronic filing *medium* (a technology issue) complacently assumes the existence of a filing *system* (a management one).

Secondly, there is the assumption that the digital format can capture everything. Contextual information can be provided to augment the understanding of a particular real-world artefact, but it will sometimes fall short of need or expectation.

Thirdly, there is the assumption that ‘the library can always sort it all out later’. Leaving information management to the end of the road means that much potentially helpful information is not ‘captured’ until it is too late.

Food for thought

A comment about context

Digital media cannot capture everything.

An HTML text rendition on a Web page of a real-life account of battle is no substitute for seeing a bloodstained original letter.

Similarly, there could be no digital parallel of the ‘vinegar search engine’ described by Paul Duguid in *The Social Life of Information*, or a suitable description of the import of the eleven words written by Richard Nixon when tendering his resignation as President of the United States.

These examples show that the full significance of a non-electronic original can have as much to do with *context* as with *content*. The electronic versions of such originals are but information world ‘surrogates’ of the real-world artefacts.

Text is not everything.

In a situation of digital rot, when the electronic file goes missing, everyone looks back for the paper copies, still considered as *the* reference format. Paper is still often the defining medium in corporate information management culture: unless ‘it’ is on paper (or at least, *also* on paper), a document or information is often still not taken seriously. Accounts departments notoriously often maintain that they need paper receipts or proofs of purchase. One can attach to a sales invoice a copy of the artwork that your enterprise has paid for, but how does one do the same for an on-screen animation?

One reason for this, and the more general obsession with the paper format, is that a document on paper is considered as being committed irreversibly. Whereas an electronic data file can be modified, a Web page may change, paper is a 'terminal format' – once printed it is considered immutable, even if it might yellow and fade a little.

Multiplicity of formats and media

Electronic filing systems have rapidly evolved beyond being mere digital copies of their paper equivalents. The multiplicity of formats, whether:

- different formats of the same content, such as a Web page, a WAP phone or a word-processor file), or
- radically different content, such as a word processor file, a movie, an architectural plan or a workflow chart,

share the common denominator of being able to be stored on the same digital medium, magnetic or optical.

The 'all digital' approach would seem therefore to answer a librarian or documentalist's prayer to dispose of or replace the multitude of storage systems necessary for media as diverse as thirteenth-century illuminated manuscript and newsreel footage.

Food for thought

'Keep a copy'

In one Central European state, the archives service – facing serious shortage of storage space – was instructed to weed out those papers and documents that could be safely disposed of or digitized.

Presented with a recommendation to throw out hundreds of thousands of paper files and thus free up valuable shelf space, the service head gave his go-ahead with one awkward proviso: 'Photocopy everything first.'

Apocryphal? Maybe. Believable? Certainly.

Furthermore, each medium and each content type has seen the evolution of specific cataloguing systems in response to their particular needs, from the humble record card for a book collection to more complete reference information, abstracts and keywords to identify and describe the catalogued item. Insufficient effort seems to have been dedicated however to developing cataloguing and description systems independently of medium of format, with consequent duplication of effort and, worse, incompatible information across different formats.

A further problem related to formats is concerned with the management of *similar* information in *dissimilar* – and often incompatible – data formats.

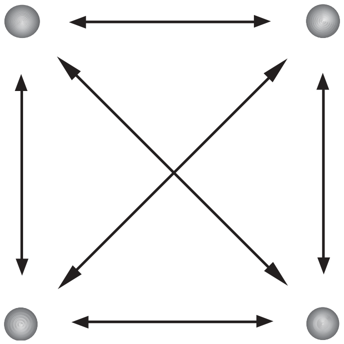

Figure 1. With only four ‘nodes’, there are already twelve connectors…

With four different systems, we can see that there are already twelve ‘interfaces’ – six bi-directional connectors between each pair of systems. If information is to flow between all four, and all four maintain their respective content in four different formats, a different ‘translation’ of that information is required for each of the twelve interfaces.

The number of interfaces increases exponentially with the number of systems added to the ‘matrix’:

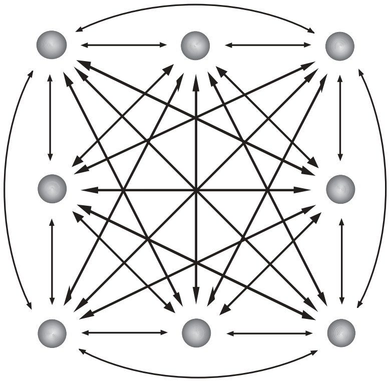

Figure 2. …and with more, the situation becomes rapidly unmanageable